



# REGRESSION ANALYSIS

## REGRESSION

- Regression Analysis measures the nature and extent of the relationship between two or more variables, thus enables us to make predictions.
- Regression is the measure of the average relationship between two or more variables.



## UTILITY OF REGRESSION

- Degree & Nature of relationship
- Estimation of relationship
- Prediction
- Useful in Economic & Business Research



# DIFFERENCE BETWEEN CORRELATION & REGRESSION

- Degree & Nature of Relationship
  - Correlation is a measure of degree of relationship between X & Y
  - Regression studies the nature of relationship between the variables so that one may be able to predict the value of one variable on the basis of another.
- Cause & Effect Relationship
  - Correlation does not always assume cause and effect relationship between two variables.
  - Regression clearly expresses the cause and effect relationship between two variables. The independent variable is the cause and dependent variable is effect.



# DIFFERENCE BETWEEN CORRELATION & REGRESSION

## ○ Prediction

- Correlation doesn't help in making predictions
- Regression enable us to make predictions using regression line

## ○ Symmetric

- Correlation coefficients are symmetrical i.e.  $r_{xy} = r_{yx}$ .
- Regression coefficients are not symmetrical i.e.  $b_{xy} \neq b_{yx}$ .

## ○ Origin & Scale

- Correlation is independent of the change of origin and scale
- Regression coefficient is independent of change of origin but not of scale

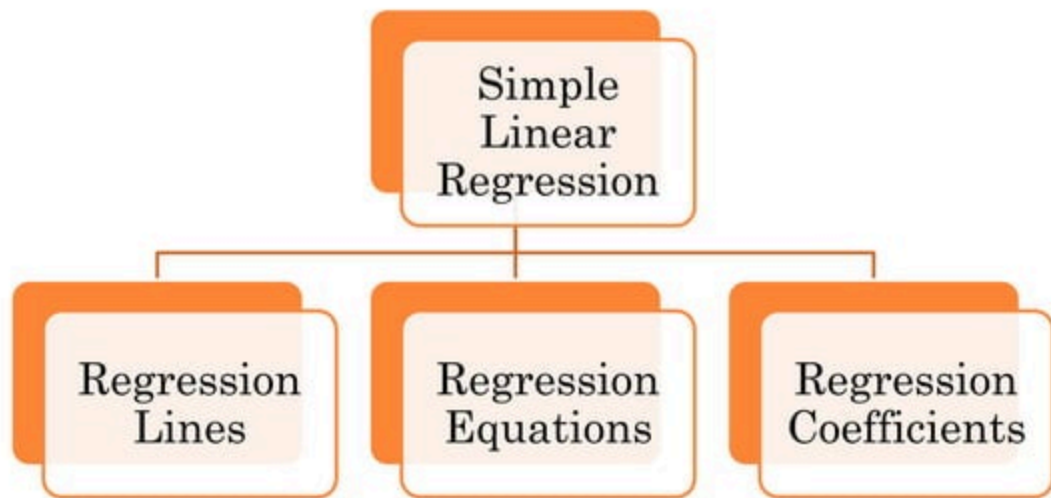


## TYPES OF REGRESSION ANALYSIS

- Simple & Multiple Regression
- Linear & Non Linear Regression
- Partial & Total Regression



# SIMPLE LINEAR REGRESSION



# REGRESSION LINES

- The regression line shows the average relationship between two variables. It is also called Line of Best Fit.
- If two variables X & Y are given, then there are two regression lines:
  - Regression Line of X on Y
  - Regression Line of Y on X
- Nature of Regression Lines
  - If  $r = \pm 1$ , then the two regression lines are coincident.
  - If  $r = 0$ , then the two regression lines intersect each other at  $90^\circ$ .
  - The nearer the regression lines are to each other, the greater will be the degree of correlation.
  - If regression lines rise from left to right upward, then correlation is positive.





# REGRESSION EQUATIONS

- Regression Equations are the algebraic formulation of regression lines.
- There are two regression equations:
  - Regression Equation of Y on X
    - $Y = a + bX$
    - $Y - \bar{Y} = byx (X - \bar{X})$
    - $Y - \bar{Y} = r \cdot \frac{\sigma_y}{\sigma_x} (X - \bar{X})$
  - Regression Equation of X on Y
    - $X = a + bY$
    - $X - \bar{X} = bxy (Y - \bar{Y})$
    - $X - \bar{X} = r \cdot \frac{\sigma_x}{\sigma_y} (Y - \bar{Y})$



## REGRESSION COEFFICIENTS

- Regression coefficient measures the average change in the value of one variable for a unit change in the value of another variable.
- These represent the slope of regression line
- There are two regression coefficients:
  - Regression coefficient of Y on X:  $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$
  - Regression coefficient of X on Y:  $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$

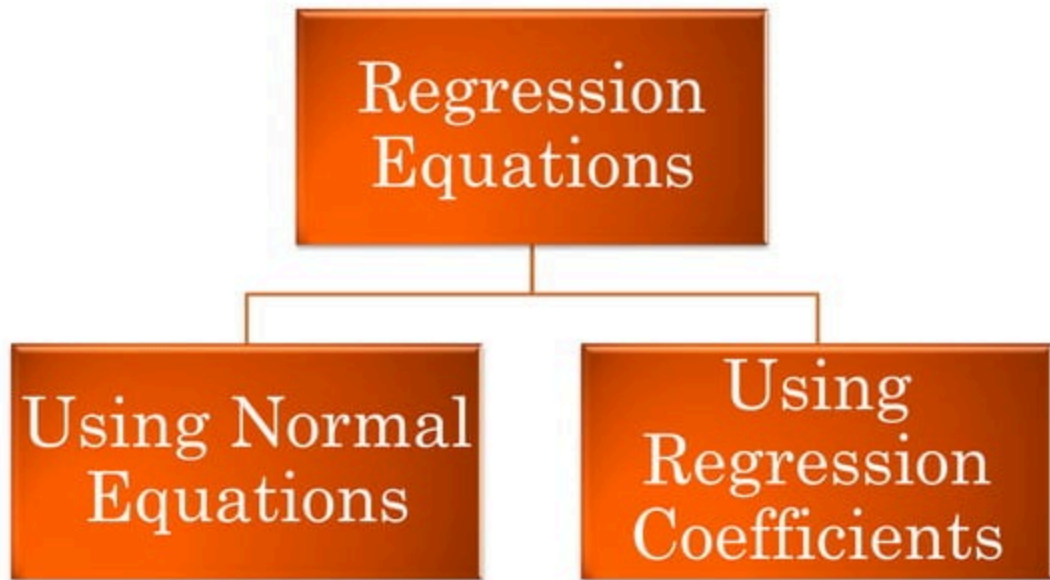


## PROPERTIES OF REGRESSION COEFFICIENTS

- Coefficient of correlation is the geometric mean of the regression coefficients. i.e.  $r = \sqrt{b_{xy} \cdot b_{yx}}$
- Both the regression coefficients must have the same algebraic sign.
- Coefficient of correlation must have the same sign as that of the regression coefficients.
- Both the regression coefficients cannot be greater than unity.
- Arithmetic mean of two regression coefficients is equal to or greater than the correlation coefficient. i.e.  $\frac{b_{xy} + b_{yx}}{2} \geq r$
- Regression coefficient is independent of change of origin but not of scale



## OBTAINING REGRESSION EQUATIONS



# REGRESSION EQUATIONS IN INDIVIDUAL SERIES USING NORMAL EQUATIONS

- This method is also called as Least Square Method.
- Under this method, regression equations can be calculated by solving two normal equations:
  - For regression equation Y on X:  $Y = a + bX$ 
    - $\Sigma Y = Na + b\Sigma X$
    - $\Sigma XY = a\Sigma X + b\Sigma X^2$
  - Another Method
    - $b_{yx} = \frac{N \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{N \cdot \Sigma X^2 - (\Sigma X)^2}$  &  $a = \bar{Y} - b\bar{X}$
- Here a is the Y – intercept, indicates the minimum value of Y for X = 0
- & b is the slope of the line, indicates the absolute increase in Y for a unit increase in X.



## PRACTICE PROBLEMS

Q1: Calculate the regression equation of X on Y using method of least squares:  $X = 0.5 + 0.5Y$

X	1	2	3	4	5
Y	2	5	3	8	7

Q2: Given the following data:

$$N = 8, \Sigma X = 21, \Sigma X^2 = 99, \Sigma Y = 4, \Sigma Y^2 = 68, \Sigma XY = 36$$

Using the values, find:

- Regression Equation of Y on X
- Regression Equation of X on Y
- Value of Y when X = 10
- Value of X when Y = 2.5

$$Y = -1.025 + 0.581X$$

$$X = 2.432 + 0.386Y$$

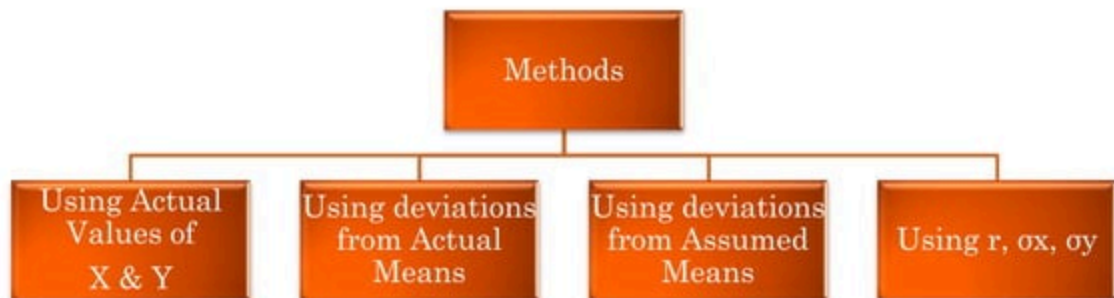
$$Y = 4.785$$

$$X = 3.397$$





# REGRESSION EQUATIONS USING REGRESSION COEFFICIENTS





## REGRESSION EQUATIONS USING REGRESSION COEFFICIENTS (USING ACTUAL VALUES)

### ○ Regression Equation of Y on X

- $Y - \bar{Y} = b_{yx} (X - \bar{X})$  where  $b_{yx} = \frac{N \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{N \cdot \Sigma X^2 - (\Sigma X)^2}$

### ○ Regression Equation of X on Y

- $X - \bar{X} = b_{xy} (Y - \bar{Y})$  where  $b_{xy} = \frac{N \cdot \Sigma XY - \Sigma X \cdot \Sigma Y}{N \cdot \Sigma Y^2 - (\Sigma Y)^2}$

Q3: Calculate the regression equation of Y on X & X on Y

$$Y = 1.3X + 1.1, X = 0.5 + 0.5Y$$

X	1	2	3	4	5
Y	2	5	3	8	7



## REGRESSION EQUATIONS USING REGRESSION COEFFICIENTS (USING DEVIATIONS FROM ACTUAL VALUES)

### ○ Regression Equation of Y on X

- $Y - \bar{Y} = b_{yx} (X - \bar{X})$  where  $b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

### ○ Regression Equation of X on Y

- $X - \bar{X} = b_{xy} (Y - \bar{Y})$  where  $b_{xy} = \frac{\Sigma xy}{\Sigma y^2}$

Q4: Calculate the regression equation of Y on X & X on Y using method of least squares:  $Y = 0.26X + 3.2$ ,  $X = 4.75 + 0.45Y$

X	2	4	6	8	10	12
Y	4	2	5	10	3	6



## REGRESSION EQUATIONS USING REGRESSION COEFFICIENTS (USING DEVIATIONS FROM ASSUMED MEAN)

### ○ Regression Equation of Y on X

- $Y - \bar{Y} = b_{yx} (X - \bar{X})$  where  $b_{yx} = \frac{N \cdot \Sigma dx dy - \Sigma dx \Sigma dy}{N \cdot \Sigma dx^2 - (\Sigma dx)^2}$

### ○ Regression Equation of X on Y

- $X - \bar{X} = b_{xy} (Y - \bar{Y})$  where  $b_{xy} = \frac{N \cdot \Sigma dx dy - \Sigma dx \Sigma dy}{N \cdot \Sigma dy^2 - (\Sigma dy)^2}$

Q5: Calculate the regression equation of Y on X & X

$$Y = 1.212 X + 34.725$$

X	78	89	97	69	59	79	68	61
Y	125	137	156	112	107	136	124	108





## REGRESSION EQUATIONS USING REGRESSION COEFFICIENTS (USING STANDARD DEVIATIONS)

### ○ Regression Equation of Y on X

- $Y - \bar{Y} = b_{yx} (X - \bar{X})$  where  $b_{yx} = r \cdot \frac{\sigma_y}{\sigma_x}$

### ○ Regression Equation of X on Y

- $X - \bar{X} = b_{xy} (Y - \bar{Y})$  where  $b_{xy} = r \cdot \frac{\sigma_x}{\sigma_y}$

Q6: Estimate Y when X = 9 as per the following information:

$$Y = 15.88$$

	X	Y
Arithmetic Mean	5	12
Standard Deviation	2.6	3.6
Correlation Coefficient	0.7	



## PRACTICE PROBLEMS

Q7: If  $\bar{X} = 25$ ,  $\bar{Y} = 120$ ,  $b_{xy} = 2$ . Estimate the value of X  
when  $Y = 130$ . X = 45

Q8: If  $\sigma_x^2 = 9$ ,  $\sigma_y^2 = 1600$ , obtain  $b_{xy}$ .  $b_{xy} = 0.04$

Q9: Given two regression equations:

$$3X + 4Y = 44$$

$$5X + 8Y = 80$$

Variance of X = 30.

Find  $\bar{X}$ ,  $\bar{Y}$ , r and  $\sigma_y$  8,5,- 0.91, 3.7



## SHORTCUT METHOD OF CHECKING REGRESSION EQUATIONS

○ Suppose two regression equations are as follows:

- $a_1x + b_1y + c_1 = 0$
- $a_2x + b_2y + c_2 = 0$

Case 1: If  $a_1b_2 \leq a_2b_1$  (in magnitude, ignoring negative), then

- $a_1x + b_1y + c_1 = 0$  is the regression of Y on X
- $a_2x + b_2y + c_2 = 0$  is the regression of X on Y

Case 2: If  $a_1b_2 > a_2b_1$  (in magnitude, ignoring negative), then

- $a_1x + b_1y + c_1 = 0$  is the regression of X on Y
- $a_2x + b_2y + c_2 = 0$  is the regression of Y on X



## STANDARD ERROR OF ESTIMATE

- Standard error of estimate helps us to know that to what extent the estimates are accurate.
- It shows that to what extent the estimated values by regression line are closer to actual values
- For two regression lines, there are two standard error of estimates:
  - Standard error of estimate of Y on X ( $S_{yx}$ )
  - Standard error of estimate of X on Y ( $S_{xy}$ )





## FORMULAE FOR SE (Y ON X)

- $S_{yx} = \sqrt{\frac{\Sigma(Y - Y_c)^2}{N}}$       Y = Actual Values,  
Y<sub>c</sub> = Estimated Values
- $S_{yx} = \sqrt{\frac{\Sigma Y^2 - a\Sigma Y - b\Sigma XY}{N}}$       Here a & b are to be  
obtained from normal equations
- $S_{yx} = \sigma_y \sqrt{1 - r^2}$



## PRACTICE PROBLEMS – SE

Q10: Find the Standard error of estimates if  $\sigma_x = 4.4$ ,  
 $\sigma_y = 2.2$  &  $r = 0.8$                       Ans: 1.32, 2.64

Q11: Given:  $\Sigma X = 15$ ,  $\Sigma Y = 110$ ,  $\Sigma XY = 400$ ,  $\Sigma X^2 = 250$ ,  
 $\Sigma Y^2 = 3200$ ,  $N = 10$ . Calculate  $S_{yx}$                       Ans: 13.21

Q12: Compute regression equation Y on X. Hence, find  $S_{yx}$   
Ans:  $Y = 11.9 - 0.65X$ , 0.79

X	6	2	10	4	8
Y	9	11	5	8	7

